

# DATOS ABIERTOS DE INVESTIGACIÓN. CAMINO RECORRIDO Y CUESTIONES PENDIENTES

**Fernanda Peset Mancebo\***

Universidad Politécnica de Valencia

**Rafael Aleixandre-Benavent**

CSIC Ingenio-Universidad de Valencia

**Yolanda Blasco-Gil**

Universidad de Valencia

**Antonia Ferrer-Sapena**

Universidad Politécnica de Valencia

**Resumen:** Los aspectos profesionales y técnicos de la gestión de datos de investigación han avanzado, por ejemplo las infraestructuras tecnológicas o los esquemas de metadatos de descripción. Sin embargo, otras cuestiones siguen necesitando ser investigadas, y más ahora en que los datos han de estar disponibles en abierto. Este trabajo efectúa una revisión bibliográfica y de sitios web de referencia que permite identificar los últimos avances producidos y las tendencias de investigación en temas relacionados con los datos científicos. La publicación de datos en revistas, el estudio del comportamiento de los autores o los métodos para medir el impacto de los datos son aspectos que necesitan ser abordados con urgencia. El objetivo de este trabajo es resaltar los retos pendientes en este momento en que se está produciendo un cambio disruptivo en su intercambio, dada la cantidad de datos disponibles en abierto, reflexión que nos permitimos desde nuestra experiencia en este campo.

**Palabras clave:** Datos de investigación; datos abiertos; métodos de investigación; investigación científica.

**Title:** OPEN RESEARCH DATA. THE TRAVELLED WAY AND PENDING ISSUES.

**Abstract:** The professional and technical aspects associated with the management of research data have advanced: technical infrastructures or descriptive metadata schemes. There are others, related with the research field, which are still under study due to the need now to make them available as open data. This paper makes a bibliographical review and reference web sites that allow to identify the latest advances and research trends on issues related with research data. Some of them need to be addressed urgently in our opinion and experience in this field: data publication in journals, study of the behaviour of authors and methods to measure the impact of the data. The aim of this work is to highlight the outstanding challenges in a moment when there is a disruptive change in its exchange given the amount of data available as open data.

**Keywords:** Research data; open data; research methods; scientific research.

## 1 INTRODUCCIÓN

La era de los datos ya está aquí: en 2017 se producirán 7,7 zettabytes de datos según una previsión de Cisco (2013). Ante la avalancha de información, la sociedad debe hacer frente a los retos de almacenarla, analizarla y comunicarla. Surgen oportunidades para crear negocio, crear conocimiento y se buscan los métodos para convertirlos en evidencias. Las nuevas tareas y profesiones pivotan alrededor de temas como:

- Periodismo de datos.
- Participación ciudadana en los gobiernos.
- Transparencia.

- Extracción de conocimiento con técnicas, por ejemplo, de *data mining* sobre la web social, los repositorios o las bases de datos.
- Reutilización de datos, especialmente los del sector público.

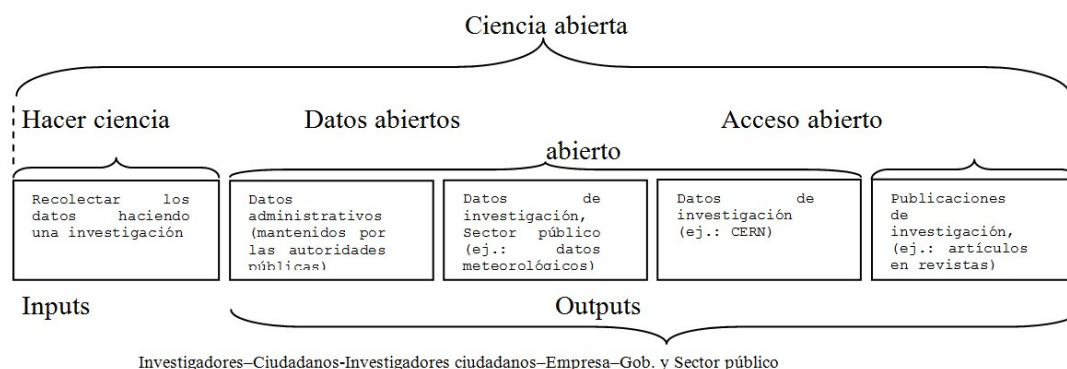
El análisis de evidencias y datos para interpretar la realidad pretende ser base irrefutable de estos nuevos conocimientos y servicios mencionados. Utilizar los datos como pruebas aproxima este nuevo modelo al proceder de los científicos, pero con una diferencia sustancial. El énfasis en la reutilización de datos obliga a explorar nuevas

\* mpesetm@upv.es

Recibido: 19-10-2016; 2ª versión: 30-01-2017; 3ª versión: 10-02-2017; aceptado: 15-02-2017.

PESET MANCEBO, F.; ALEIXANDRE BENAVENT, R.; BLASCO GIL, Y. y FERRER SAPENA, A. Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. *Anales de Documentación*, 2017, vol. 20, nº 1. Disponible en: <http://dx.doi.org/10.6018/analesdoc.20.1.272101>.

formas de trabajo, de innovación social y de negocio. Ahora se busca integrar datos procedentes de un gran número de bases de datos de forma dinámica y actualizados automáticamente. Para crear conocimiento se usan técnicas potentes de interrogación y *data mining*, que permiten plantear interrogantes hasta ahora impensables, especialmente si hablamos de datos masivos (Boulton, 2013). A este cambio, que puede compararse con la revolución de Gutenberg o la que se produce en el siglo XVIII con la aparición de las revistas, se suma su disponibilidad en abierto. Podemos afirmar que por la cantidad de datos y por su disponibilidad, se está propiciando un verdadero cambio de paradigma que se basa en lo abierto. En palabras de Boulton (2015) «La explotación efectiva de datos masivos depende fundamentalmente de una cultura internacional de “datos abiertos”». Cultura que responde a un ecosistema complejo, como refleja la Figura 1



**Figura 1. Open science (Boulton, 2013).** <http://goo.gl/eYY1Ls>.

Tanto la *Agenda europea sobre ciencia abierta* (Comisión Europea, 2016) como Piwowar *et al.* (2008) señalan la necesidad de modificar los hábitos y comportamientos de investigadores y organizaciones ante lo abierto. Para responder a este avance se necesitan soluciones sociotécnicas que den sentido a las infraestructuras puramente tecnológicas. Al contrario que los cambios técnicos, los avances sociales no se producen con rapidez. A ello se suma en el caso de los datos abiertos de investigación que la evolución se produce a un tiempo en la arena internacional y en los diferentes países, lo que exige una acción combinada. «La ciencia es una actividad internacional, ejecutada en un marco cultural nacional, requiriendo así estrategias nacionales que encajen en un marco internacional común» (Boulton, 2015).

## 2 ANTECEDENTES

Esta naturaleza multidimensional del trabajo con datos es la que complica su gestión (Couto-Corrêa, 2016). Si los datos de investigación además se disponen en abierto, entran a formar parte del contexto más amplio de la *open science* y por extensión del movimiento *open data* ligado a la rendición de cuentas y la reutilización (Ferrer-Sapena; Peset; Aleixandre-Benavent, 2011; Ferrer-Sapena; Sánchez-Pérez, 2013; Peset; Fernández-López, 2014; Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero, 2014). De hecho, la cumbre de líderes del G8 de junio de 2013 califica los datos derivados de la investigación y los resultados de experimentos científicos como especialmente valiosos para la economía (Unión Europea, 2014); algo que retoma Moedas (2015) al definirlos como objetivos estratégicos de la Unión Europea para la innovación abierta, la ciencia abierta y la apertura al mundo.

En el nivel de política científica, observamos que los requerimientos para liberar los datos se han consolidado, lo que ha acabado influyendo en otros ámbitos como puede ser su propia definición. Los datos obtenidos con fondos públicos están sujetos a políticas de liberación específicas, como parte que son de la información del sector público. En Estados Unidos, el primer mandato de los National Institutes of Health (NIH) en 2003 se recogió en las instrucciones para la financiación de proyectos de la National Science Foundation (2010), que «espera que los investigadores compartan los datos primarios y otros materiales creados». En 2013 esta tendencia se ve reforzada por la Fair Access to Science and Technology Research Act (FASTR). Provocó que el director de la U.S. Office of Science and Technology Policy (OSTP) instase a todas las agencias federales de financiación públicas para que publicaran su política de acceso abierto a los resultados de investigación (publicaciones y datos) (Halbert, 2013; Berman; Cerf, 2013).

Por su parte, la Unión Europea cristalizó su deseo de hacer públicos los datos en la primera convocatoria del programa *H2020*. Los socios que participan en el *Open Research Data Pilot* se obligan a liberar «Los datos necesarios para validar las publicaciones de investigación, con sus metadatos, tan pronto como sea posible; y otros datos, con sus

metadatos, dentro del plazo previsto en sus planes de gestión de datos así como el momento de liberación que indicaron», según el artículo 29.3 del *Agreement* (Unión Europea, 2016a). Esta iniciativa piloto ha alcanzado tal calado que para 2017 se ha convertido en el mantra “open as default” (Unión Europea, 2016b). En España, la *Ley 14/2011 de la ciencia, la tecnología y la innovación* (España, 2011) no incluía referencia a los datos. Ahora bien, la convocatoria de proyectos del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica –la última de marzo 2016– los relaciona con la *Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público* (España, 2015). Proporciona indicaciones y obligaciones para las áreas de estudios oceanográficos, polares y de ciencias sociales.

En cuanto al estatuto ontológico de los datos de investigación existen varias definiciones. La que más consenso aúna al haber sido adoptada por entidades como los National Institutes of Health (NIH, 2003) de Estados Unidos o la OECD (2007) es la que considera datos de investigación todo aquel material que sirve para certificar los resultados de la investigación que se realiza, que se han registrado durante ella y que se ha reconocido por la comunidad científica (Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012). En principio no hace referencia a su apertura. Para ello, la Royal Society (2012) define los datos abiertos como aquéllos que cumplen ciertos criterios cuando son liberados, lo que se llama apertura inteligente/competitiva: son accesibles, usables, evaluados/certificados e inteligibles; aspectos que recogerán con posterioridad los principios FAIR (findables, accesibles, interoperables, reusables) de Force11 (Doorn; Dillo, 2016) y que Méndez propone ampliar con dos nuevas «r», FAIRrr (2017): fiables y reproducibles (Méndez, 2016). Esto significa que hacerlos disponibles, liberarlos, no los incluye automáticamente en la órbita de lo abierto. La descripción asociada es la que permitirá que puedan ser reutilizados y, en consecuencia, indicio de la calidad de un dato de investigación abierto.

Es esta apertura la que establece la diferencia en estos momentos, ya que el intercambio de datos de investigación y la colaboración en ciencia ha sido habitual. La solicitud de datos o publicaciones a su autor es una práctica que incluso se incorpora en los sistemas como Research Gate (Porter, 2010; Torres-Salinas, 2010). Ahora muchos datos de investigación se están liberando en repositorios para una comunidad de reutilizadores desconocida a priori, por lo que precisa de una descripción correcta que sustituya el contacto con el productor original de esos datos.

Como apuntábamos, estas definiciones tienen una utilidad relativa cuando un investigador se enfrenta al momento de liberar sus datos, quien en la práctica se limita a cumplir con los requerimientos para la financiación. Por esta razón insistimos en que los mandatos influyen sobre la definición y naturaleza de los datos liberados. A efectos prácticos, puede afirmarse que los datos de investigación abiertos se definen como los necesarios para validar una investigación financiada públicamente, es decir sus evidencias.

Otros aspectos, especialmente en el ámbito profesional, han avanzado también (Elger *et al.*, 2016). Contamos ya con casos consolidados que demuestran que, al menos fuera de nuestro país, la gestión de este tipo de información está encarrilada. Estas iniciativas pueden proceder de la propia comunidad de práctica, como Research Data Alliance (RDA), o de los niveles más altos de coordinación de la investigación, como el Comité de Datos del International Council for Science, Codata. Se trata de programas orientados a ofrecer servicios desde el sector infomediario, en el que se encuentran las bibliotecas, lo que supone que su orientación sea más práctica que de investigación. A modo de ejemplo cabe citar el Data Curation Centre, que ya hace años abordó la conservación y preservación de los datos de investigación. Las infraestructuras tecnológicas son también numerosas, como Pangea, Zenodo o Dryad. Aparecen recogidas en el directorio mundial re3data.org, que informa de 1.500 en abril de 2016 (re3data, 2016; Pampel *et al.*, 2013; González *et al.*, 2013), y en Data Citation Index (DCI) de Thomson Reuters, que pasó de seleccionar 69 repositorios en 2013 (Peset; González, 2017) a más de 300 según estimamos en mayo de 2016. Otros aspectos técnicos también ilustran que en el terreno profesional la gestión de datos está asentada. Desde 1995, el esquema de metadatos Data Documentation Initiative (DDI) es capaz de reflejar el ciclo de vida de los datos (Martínez-Uribe; Macdonald, 2008; Vlaeminck; Wagner, 2014). También Dublin Core extendido o el esquema DataCite (2014) se encuentran ampliamente respaldados (Force; Robinson, 2014). Un último aspecto para su control, la elección de un sistema de identificador único, también se ha zanjado a favor de asignar DOI, incluso en Dataverse donde previamente se utilizaba handle.

La hipótesis de este trabajo es que aún subsisten algunos aspectos que necesitan ser desarrollados en este nuevo escenario de lo abierto. Piwowar y Chapman (2010) detectan que la investigación sobre cómo se publican, consumen y reutilizan los datos no se ha acabado de definir, aunque se cuente con avances. Hay cuestiones que se investigan de manera parcial, e incluso algunas, las menos relacionadas con la práctica profesional, están aún en proceso de definición, como:

- La evolución de los modelos de publicación de datos;

- El estudio del comportamiento de los autores respecto a la producción y reutilización;
- Y las métricas de uso.

Este trabajo efectúa una revisión bibliográfica y de sitios web de referencia para identificar los últimos avances producidos y las tendencias de investigación en temas que necesitan ser abordados con urgencia. Su objetivo es resaltar los retos pendientes en este momento en que se está produciendo un cambio disruptivo en su intercambio, dada la cantidad de datos disponibles en abierto.

### 3 PUBLICACIÓN DE LOS DATOS

Los datos se están haciendo públicos de dos formas básicas, que coinciden con las vías verde y dorada del acceso abierto (Nina-Alcocer; Blasco-Gil; Peset, 2013). Esto involucra a diferentes actores, cada uno con sus propios intereses, y supone diversos niveles de apertura, ya que no todo lo libre coincide con lo abierto (Peset; González, 2017).

La vía verde es la de los repositorios, como el disciplinar Protein Data Bank, el generalista Dryad o los institucionales. Se encuentran vinculados, respectivamente, a su comunidad científica, a los gestores de información y empresas, y finalmente a las instituciones. Esta vía está bastante explorada. Cuenta con una amplia tradición de interoperabilidad y de servicios de búsqueda de valor añadido cuya nota diferenciadora es su interés en garantizar la preservación. En ella quedan dos situaciones aún en estudio. En primer lugar cómo ofrecer estadísticas normalizadas (Doorn y Dillo, 2016), que analizaremos en el último apartado de métricas. En segundo lugar, que no está decidido cómo vincular la publicación y los datos que subyacen a ella, así como definir la relación que existe entre datos de investigación que están directamente vinculados a una publicación. El camino cuando hay una publicación ya está iniciado: las revistas recomiendan ciertos depósitos de datos y obligan a incluir su identificador en el artículo. Este es el modelo de negocio de Dryad y el consorcio de editoriales y asociaciones científicas que lo conforman. Por su parte, otros sistemas, como Figshare, han comenzado a ofrecer servicios en forma de repositorios de las propias revistas, incluso Dataverse vinculada a una herramienta tan conocida como Open Journal System.

La vía dorada supone el almacenamiento los datos de investigación como material suplementario junto al artículo en la plataforma de las editoriales o bien la publicación de *data papers*. Da lugar a nuevos modelos de publicación:

- Las *data journals* publican *data papers* (Chavan; Penev, 2011, García-García; López-Borrull; Peset, 2015).
- Las *megajournals* como *Plos ONE* aceptan artículos con resultados preliminares (Björk; Catani, 2016).
- Títulos clásicos como *Nature* o *Cell* están aceptando material suplementario (Martinsen, 2012).

Esta vía tiene implicaciones a varios niveles. El hecho de depositar los datos junto al artículo como material complementario es, sin duda, un procedimiento muy apreciado por los investigadores (Dallmeier-Tiessen *et al.*, 2012), ya que el ecosistema científico gira alrededor de la publicación en revistas. Las editoriales científicas están adaptándose a esta nueva situación, tal y como sucedió con el movimiento de acceso abierto. Los *article processing charges* crearon un nuevo modelo de negocio que ha supuesto mayores costes que la ruta verde (Johnson; Pinfield; Fosci, 2015). Aun así existen medidas para evitar que parte de la financiación de la investigación pague los gastos de publicación en revistas que en ocasiones pertenecen a empresas privadas. A ello se suma, según Filippov (2014), que la legislación europea sobre propiedad intelectual defiende los derechos de la industria editorial, lo que constituye una rémora para la explotación de los datos para fines comerciales, a diferencia de Estados Unidos o China.

En estos momentos todavía se están experimentando y estandarizando las instrucciones para el autor con respecto a los datos (Piwowar; Chapman, 2010; Aleixandre-Benavent *et al.*, 2015; CSUC, 2016). Está claro que los intereses de unas editoriales cada vez más concentradas (Larivière; Haustein; Mongeon, 2015; Ware; Mabe, 2015) no se centran en su propiedad e intercambio. No contamos hoy en día con una solución clara. Se percibe cierta indecisión, lo que sitúa el estudio de las revistas que aceptan material suplementario en el centro de un problema que se mueve entre:

- La necesidad de proporcionar el material suplementario junto al artículo en el préstamo interbibliotecario (NISO, 2013).
- La falta de interoperabilidad.
- La escasa garantía de que sean preservados en el futuro.
- El hecho de dar acceso a los datos aun no siendo suscriptor de la revista.
- Y las políticas confusas sobre qué se puede hacer con ellos y qué no (García-García; Peset, 2014).

En nuestra opinión, según qué vía tome la delantera como modelo de publicación de datos, la preservación estará en manos públicas o privadas y la interoperabilidad será o no facilitada.

#### 4 RECONOCIMIENTO, ACCESO Y REUTILIZACIÓN. COMPORTAMIENTO DE LOS AUTORES

En su actitud frente a los datos observamos dos similitudes con el proceso que los autores siguieron para dar acceso abierto a sus publicaciones:

- Su decisión de liberar datos dependerá de que las agencias de evaluación de la ciencia reconozcan este trabajo.
- Sufren una contradicción que fue ya bautizada como paradoja del Dr. Jekyll y Mr. Hyde: como investigadores desean acceder a los datos ajenos, mientras que como autores se resisten a compartirlos.

Todavía no están bastante estudiados estos fenómenos: intuimos los problemas, pero no tenemos datos suficientes para medir su grosor. Veámoslo con detenimiento. En relación al primer punto, tanto autores como organismos se ven forzados a adaptarse a criterios tradicionales de evaluación basados en patentes e impacto de las publicaciones. Por el contrario, los financiadores y otros agentes claman por hacer disponibles los datos procedentes de la investigación. Si esta labor no es reconocida en los procesos de evaluación, será muy difícil que se pueda disponer de los datos en abierto de forma generalizada y que se obtengan los potenciales beneficios que siempre se mencionan (Ball; Duke, 2015; Nosek *et al.*, 2015; Borgman, 2015; He; Nahar, 2016). Las autoridades académicas y quienes dirigen la política científica deberían considerar la publicación de datos como parte de la carrera de un investigador, pues los resultados de investigación han sufrido un cambio sustancial (Laoie; Childress *et al.* 2016). El reconocimiento tiene profundas consecuencias en el nivel técnico y en la práctica de los investigadores. Influye en la necesidad de fijar las formas de citación (Altman; Crosas, 2013), pero también en investigar los métodos e indicadores que midan los trabajos de producir datos en abierto y de consumir datos ajenos, como más adelante se explicará.

El segundo punto se refiere al comportamiento de los autores con respecto a esta producción y acceso a datos de investigación. Se llevan a cabo multitud de encuestas, pero no existe un panorama uniforme sobre ello y son difícilmente comparables entre sí (Peset *et al.*, 2015). Cabe destacar Parse.Insight para Europa (Comisión Europea, 2012), Datares para Norteamérica (Keralis *et al.*, 2013) o una para Austria que esboza con maestría los interrogantes que preocupan a los autores (Bauer *et al.*, 2015). En nuestro país todavía no contamos con una a escala nacional, aunque sí para instituciones individuales (Ferrer-Sapena; Aleixandre-Benavent, 2016; Bernal; Román, 2014) y otras ejecutadas desde organizaciones como CSUC (Alcalá; Anglada, 2016) y previstas por Rebiun <<http://blogs.uji.es/rebiunworkshop>>.

En general, todas reflejan un índice bajo a la hora de producir datos en abierto, así como las reticencias para compartirlos incluso en una de las áreas más proclives, la biomédica (Piwowar; Chapman, 2008; Rani; Buckley, 2012; Borgman, 2013; Fecher *et al.*, 2015). Por supuesto, las diferencias entre disciplinas son acusadas; tanto que Borgman (2015) distingue entre *big data* y *little data* o la larga cola de pequeños *datasets* que conforma gran parte del entramado científico.

Las carencias en el descubrimiento y acceso a los datos susceptibles de ser reutilizados se producen a escala mundial. La dificultad para consumir datos se agudiza por la dispersión de las fuentes y por su distinta naturaleza (Wallis; Rolando; Borgman, 2013; Collins, 2011). Nos referimos a fuentes de datos que en muchas ocasiones no se derivan del trabajo científico. Se trata de información de tipo transversal a múltiples disciplinas que se originan por parte del sector público (PSI) o por parte de una comunidad de usuarios, tal y como se apuntaba en la Figura 1 y muestra la Tabla I. El éxito en la reutilización de estas fuentes dependerá de la fiabilidad del productor de los datos, del formato en que se ofrezcan y de la posibilidad de aplicar técnicas de análisis.

Naturaleza	Origen	Ejemplos
Estadística	Agencia Nacional de Estadística/ Datos abiertos de gobiernos/ Organismos internacionales	INE; datos.gob.es; FMI o WorldBank
Geoposicionamiento	Mapas abiertos	StreetMap
Meteorología	Agencias nacionales	Aemet
Registros bibliográficos	De revistas o artículos	SJR o DOAJ
Ontologías	Vocabularios/Thesaurus/Clasificaciones	LC, Agrovoc o Getty

**Tabla I. Fuentes de datos.**

Además, para descubrirlas no hay un único directorio. En el campo de los datos de investigación los inventarios identifican depósitos de datos de otros investigadores, y no conocemos hasta el momento ninguno que reúna los que proceden del sector PSI que sean de interés para la ciencia, lo que puede ser un reto a asumir. Los inventarios responden a las vías por las que se liberan los datos de investigación. Re3data.org, que se fundió con Databib en 2013 con apoyo de DataCite, está recogiendo los repositorios (Pampel *et al.*, 2013), mientras que ODiSEA, del proyecto

Datasea, recopila las revistas que aceptan material suplementario (García-García *et al.*, 2012) como parte de nuestro cosechador de datos de investigación HODS/rd (Albiñana; Peset, 2017).

## 5 MÉTRICAS DE USO

En el ámbito científico, el reconocimiento y la atribución es crucial (Nosek *et al.*, 2015), lo que se ha resuelto mediante la fijación de la citación y de la trazabilidad del origen de la idea (Farquhar; Brase, 2014). Como consecuencia se observa la insistencia por ejemplo de DataCite en asignar DOI y de buscar la forma de citar correctamente los datos que se consumen. Es una muestra de que se sigue midiendo el impacto de un trabajo por la reutilización que tiene, concretamente por el número de citas que recibe (Ball; Duke, 2015). No obstante, en un escenario tan complejo, con datos publicados e inéditos dispersos en múltiples fuentes, el papel de las métricas alternativas se perfila como crucial (Piwowar; Vision, 2013).

Establecer un sistema métrico sólido, DataMetrics, contribuirá a evaluar el resultado del esfuerzo del investigador, con una visión más amplia de los logros y de la reputación en el ecosistema científico, y a crear nuevos incentivos que apoyen el intercambio de datos (Ferrer-Sapena *et al.*, 2016). Los métodos para poder medir el acceso y consumo de datos están en fase experimental –citación, *tracking*, *altmétricas*..., si bien se prevé que la producción de nuevo conocimiento mediante la reutilización de datos procedentes de otros investigadores u organismos sea clave en un futuro próximo (Murray–Rust; Molloy; Cabell, 2014). Existen grandes proyectos para fomentar las métricas basadas en datos, como DataLevelMetrics (University of California \$299,964.00) (Peters *et al.*, 2016) u OpenAire2020 (H2020, 13.132.500 euros), pero aún no tenemos constancia de un sistema suficientemente extendido. De hecho, proponer uno no es trivial y requiere del análisis preliminar de varios condicionantes (Costas *et al.*, 2013):

- ¿Qué métricas es posible recoger de manera automática y que sean válidas y útiles para la comunidad?
- ¿De qué fuentes de información asociadas se extraerán los datos?
- ¿Cuáles son sus limitaciones?
- ¿Cuáles son los requisitos de la comunidad científica para validar los *DataMetrics* y su fiabilidad?
- ¿Son aplicables a todas las áreas científicas?
- ¿Cuál es su sostenibilidad a largo plazo?
- ¿Permite promover la normalización de los datos, compararlos y cruzarlos?
- ¿Qué lagunas existen en las infraestructuras para que los DataMetrics se integren plenamente en el sistema de investigación (repositorios, financiadores, investigadores, centros de información)?

Para comenzar a diseñar un sistema se necesitan experimentos sobre bases muy concretas que sean capaces de interrelacionar las fuentes de datos que existen ahora. Como muestra, recogemos las formas utilizadas por varios autores para estimar la citación y reutilización de datos, aunque como hemos dicho ha de consolidarse un método más uniforme. Piwowar, con otros autores, estudiaron la disponibilidad y las citas de los datos de los *microarrays* o biosensores para la expresión de genes (Piwowar; Day; Fridsma, 2007; Piwowar; Vision, 2013). En 2013, para una muestra de unos 10.000 artículos, evidenciaron el aumento de citas recibidas en Scopus cuando los trabajos habían liberado los datos. Para la reutilización estudiaron la mención del identificador del dataSet de Gene Expression Omnibus (GEO) en artículos de PubMed Central y extrapolaron los resultados a todo PubMed.

Por su parte, Moss (2013) mostró la construcción de Bibliography of Data-Related Literature. Se trata de una lista de 60.000 registros que citan datos de ICPSR (Interuniversity Consortium for Political and Social Research) elaborada de forma manual mediante envíos de autores, *Google alerts*, cosechamiento de DOI, técnicas de *textmining*...

Robinson-García; Jiménez-Contreras; Torres-Salinas (2015) estudiaron los hábitos de citación por áreas disciplinares. Para ello, introdujeron en una base de datos relacional todos los registros del Data Citation Index (DCI) y los clasificaron en cuatro grandes áreas, coincidentes con las de los repositorios.

Peters *et al.* (2016) compararon la cobertura social de los datos de investigación, para lo que descargaron los datos de DCI desde 1960 a 2014. Seleccionaron los que tenían más de 2 citas para contrastar sus DOI o URL con tres canales de social media: ImpactStory, Altmetrics.com y PlumX.

Por último, He y Nahar, (2016) recuperaron en Scopus los *url* citados de Dryad, bajando posteriormente del repositorio los metadatos y los textos completos de las publicaciones citantes. Al asignar las categorías temáticas de Scopus y el tipo de datos (experimentales, observacionales...) pudieron representar los resultados de forma segmentada.

Existen por tanto métodos muy variados para medir la reutilización de los datos. Para cubrir los aspectos que quedan por aclarar, los experimentos que se diseñen pueden fijar la atención en los siguientes puntos o imaginar otros problemas a estudiar, como determinar el porcentaje de revistas que recogen material suplementario que no están presentes en DCI:

- Contrastar el uso de los *datasets* depositados en bancos de datos con bases de datos bibliográficas o con las propias revistas, tal y como hacen algunos de los autores revisados.
- Estudiar las métricas que están ofreciendo los repositorios, como mínimo visitas y descargas como muestra la figura para Figshare.



**Figura 2. Métricas de un *dataset* depositado en Figshare.**

<[https://figshare.com/articles/JCR\\_Journals\\_sorted\\_by\\_Impact\\_Factor\\_2011\\_with\\_the\\_JCR\\_edition\\_indication/714141](https://figshare.com/articles/JCR_Journals_sorted_by_Impact_Factor_2011_with_the_JCR_edition_indication/714141)>.

Un problema en la medición del impacto es que las fuentes para la búsqueda de datos y el cómputo de su citación son dispersas. Existen varias, con muy diferentes propósitos, que informan sobre el impacto de los datos. Para conocer la influencia de un autor puede utilizarse ImpactStory (Piwowar; Vision, 2013). Para establecer la radiografía de una institución puede utilizarse PlumX. Mientras que para una comunidad disciplinar será mejor estudiar sus propios repositorios, como ICPSR (Fear, 2013; Moss, 2013) o Global Biodiversity Information Facility (GBIF) (Konkiel, 2013).

Por el momento, de una forma integral solamente lo hace DCI. Es una base de datos multidisciplinar que indexa los principales repositorios de datos. Asocia a los estudios de datos y *datasets* que ha cosechado la información de los registros bibliográficos que los citan en WoS (Torres-Salinas; Martín-Martín; Fuente-Gutiérrez, 2014). Aunque la idea es realmente ingeniosa, cuenta con limitaciones como:

- Estar constreñida al universo de artículos recogidos en WoS.
- No recoger todos los bancos que existen: 304 bancos en abril 2016, 94 en 2014 (Robinson-García; Jiménez-Contreras; Torres-Salinas, 2015), y 68 en 2013.
- No incluir una de las formas más extendidas de publicación de datos a la que ya hemos aludido: la publicación junto al artículo.
- Tener un retraso significativo debido a que los datos citados dependen de la publicación del artículo.

En definitiva, las métricas sobre uso de los datos han de estandarizarse, lo que comienza a denominarse *datametrics* cuyo mayor reto es fundamentarlas en un conjunto multidimensional de indicadores sobre presencia, alcance, uso y repercusión de los datos de investigación.

## 7 LECCIONES APRENDIDAS

En el camino de hacer localizables, accesibles, interoperables y reutilizables los datos (FAIR), las cuestiones puramente técnicas son sólo el primer paso de la cadena de valor de la ciencia. Como señalaron hace años Codata y la National Science Foundation (NSF), “el valor de los datos reside en su uso” (National Research Council, 1997). Hemos ido observando a lo largo de este artículo que la ciencia es un ecosistema social en el cual el cúmulo de intereses en ocasiones se contraponen. Las tensiones entre los principales protagonistas en la arena de los datos de investigación (autores, editores, financiadores...) necesitan de la cooperación en varias dimensiones (política, tecnológica, orgánica o lo que se llama *science facilities*) para desbloquear la capacidad de generar valor de los datos (Thomson Reuters, 2013; Halbert, 2013; Erway, 2013).

Los mandatos institucionales o de los financiadores sobre la obligación de liberar datos son clave para que los datos de investigación se dispongan en abierto de forma generalizada. De hecho, las zonas en las que los gobiernos han decidido convertir los datos de investigación en un activo estratégico están liderando el avance a nivel mundial, cada una con su propia idiosincrasia:

- Australia, que cuenta con un sistema nacional.
- Estados Unidos, basada en una cultura de la financiación muy sólida.

- Gran Bretaña en la que existe un organismo de coordinación de servicios innovadores basados en datos (JISC) y una fuerte coordinación entre todos los agentes.

Los retos que quedan por esclarecer hacen referencia a la publicación, al comportamiento de los autores en cuanto a producción y reutilización de los datos y a los métodos para medir estos fenómenos. En primer lugar, ha de investigarse las recomendaciones de los editores de revistas sobre los datos, puesto que forman un escenario mucho más disperso que el de las políticas institucionales, las que además ya han sido recogidas en *Recode*, *Pasteur4* u *Openaire2020*. Aún con esta recopilación de mandatos todavía quedan cuestiones pendientes. La Unión Europea ha formado el grupo de expertos de alto nivel para decidir cuál aplicar, mientras que en España nuestra ley de la ciencia necesitaría ser modificada. En segundo lugar, faltaría estudiar los factores sociales del ecosistema científico. Las encuestas sobre cómo perciben los datos y las prácticas de sus propios productores esclarecerían la forma de fomentar su liberación y consumo. Por último, como hemos visto se necesita aportar métodos innovadores que puedan iluminar cómo medir la publicación y la reutilización e impacto de los datos de investigación, de manera que pueda ser objeto de reconocimiento en los currículos de los investigadores. En definitiva, el panorama ofrece avances significativos para ciertos sectores, mientras que otros se encuentran en un momento de cambio.

## NOTA

Este trabajo recibe financiación MINECO CS02012-39632-C02-02, CSO2012-39632-C02-01, CSO2015-65594-C2-1-R y CSO2015-65594-C2-2-R 2016-2018. <<http://www.datasea.es>>.

## 8 BIBLIOGRAFÍA

- ABELLA, A.; ORTIZ-DE-URBINA-CRIADO, M. y DE-PABLOS-HEREDERO, C. Meloda, métrica para evaluar la reutilización de datos abiertos [en línea]. *El profesional de la información*, 2014, vol. 23, nº 6, p. 582-588. Disponible en: <http://dx.doi.org/10.3145/epi.2014.nov.04>.
- ALBIÑANA, R. y PESET, F. OpenDataScience HODS/rd 2012-2016: how was our harvester created? <<http://www.datasea.es/dt/index.php/resultados/buscador-opendatascience>>. [Consulta: 13 de enero de 2017].
- ALCALÁ, M. y ANGLADA, L. El suport a la gestió de dades de recerca per part de les biblioteques [en línea]. En: *Seminario Gestión de Datos de Investigación*. Barcelona: Universidad de Barcelona, 19 enero 2016. Disponible en: <<http://www.slideshare.net/maredata/1601-seminari-maredat-av3defanglada-57772441>>. [Consulta: 13 de enero de 2017].
- ALEIXANDRE-BENAVENT, R. *et al.* Disponibilidad en abierto de los artículos y de los datos brutos de investigación en las revistas pediátricas españolas. *Anales de Pediatría*, 2015, vol. 82, nº 1, p. 90-94. Disponible en: <http://dx.doi.org/10.1016/j.anpedi.2013.11.014>.
- ALTMAN, M. y CROSAS, M. The evolution of data citation: from principles to implementation [en línea]. *IAssist quarterly*, 2013, vol. 37, nº 1-4, p. 62-70. Disponible en: <<http://informatics.mit.edu/publications/evolution-data-citation-principles-implementation>>. [Consulta: 13 de enero de 2017].
- BALL, A. y DUKE, M. *How to track the impact of research data with metrics*. DCC How-to Guides [en línea]. Edinburgh: Digital Curation Centre, 2015. Disponible en: <<http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics>>. [Consulta: 13 de enero de 2017].
- BAUER, B. *et al.* *Researchers and their data. Results of an Austria survey – Report 2015. Version 1.2* [en línea]. Disponible en: <http://dx.doi.org/10.5281/zenodo.34005>.
- BERMAN, F. y CERF, V. Who will pay for public Access to research data? [en línea]. *Science*, 2013, vol. 341, nº 6146, p. 616-617. Disponible en: <http://dx.doi.org/10.1126/science.1241625>.
- BERNAL, I. y ROMÁN-MOLINA, J. *Prácticas en la gestión, difusión y preservación de datos de investigación en el CSIC* [en línea]. 2014. Disponible en: <<https://digital.csic.es/handle/10261/92404>>. [Consulta: 13 de enero de 2017].
- BJÖRK, B. y CATANI, P. Peer review in megajournals compared with traditional scholarly journals: Does it make a difference? [en línea]. *Learned publishing*, 2016, vol. 29, p. 9-12. Disponible en: <http://dx.doi.org/10.1002/leap.1007>.
- BORGMAN, C.L. Big data, little data, no data: The contested landscape of data sharing and reuse. *Trends in society and information technology seminar series*. Irvine, CA: University of California, 2013.
- BORGMAN, C.L. *Big data, little data, no data. Scholarship in the networked world*. Cambridge, Massachusetts: MIT Press, 2015.
- BOULTON, G. A revolution in open science? Open data & the role of libraries [en línea]. En: *LIBER Munich, June 2013*. Disponible en: <<http://www.slideshare.net/libereurope/boulton-gsb-presentationlibermunich>>. [Consulta: 13 de enero de 2017].



- BOULTON, G. Message from president Geoffrey Boulton [en línea]. En: *International Council for Science: Committee on Data for Science and Technology*, 2015. Disponible en: <<http://www.codata.org/message-from-president-geoffrey-boulton>>. [Consulta: 13 de enero de 2017].
- CHAVAN, V. y PENEV, L. The data paper: a mechanism to incentivize data publishing in biodiversity science [en línea]. *BMC Bioinformatics*, 2011, vol. 12, Suppl. 15:S2. Disponible en: <http://dx.doi.org/10.1186/1471-2105-12-S15-S2>.
- CISCO. Índice mundial sobre entornos de nube de Cisco: previsión y metodología, 2012–2017. San José: CISCO, 2013. Disponible en: <<http://docplayer.es/1754500-Indice-mundial-sobre-entornos-de-nube-de-cisco-prevision-y-metodologia-2012-2017.html>>. [Consulta: 20 de marzo de 2017].
- COLLINS, E. Use and impact of UK research data centres [en línea]. *International journal of digital curation*, 2011, vol. 6, n° 1, p. 20-31. Disponible en: <<http://www.ijdc.net/index.php/ijdc/article/view/160>>. [Consulta: 13 de enero de 2017].
- COMISIÓN EUROPEA. *Online survey on scientific information in the digital age* [en línea]. Luxembourg: Publications Office of the European Union, 2012. Disponible en: <http://dx.doi.org/10.2777/7549>.
- COMISIÓN EUROPEA. *Agenda europea para la ciencia abierta* [en línea], 2016. Disponible en: <[https://ec.europa.eu/research/openscience/pdf/draft\\_european\\_open\\_science\\_agenda.pdf#view=fit&pagemode=none](https://ec.europa.eu/research/openscience/pdf/draft_european_open_science_agenda.pdf#view=fit&pagemode=none)>. [Consulta: 13 de enero de 2017].
- CSUC. *Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca* [en línea], Barcelona: CSUC, 2016. Disponible en: <[http://www.recercat.cat/bitstream/handle/2072/266502/1607\\_RecomanacionsSeleccionarRepositoriDades.pdf?sequence=5](http://www.recercat.cat/bitstream/handle/2072/266502/1607_RecomanacionsSeleccionarRepositoriDades.pdf?sequence=5)>. [Consulta: 13 de enero de 2017].
- COSTAS, R. et al. *The value of research data. Metrics for datasets from a cultural and technical point of view* [en línea]. Copenhagen: Knowledge Exchange. Danish Agency for Culture, 2013. Disponible en: <[http://repository.jisc.ac.uk/6205/1/Value\\_of\\_Research\\_Data.pdf](http://repository.jisc.ac.uk/6205/1/Value_of_Research_Data.pdf)>. [Consulta: 13 de enero de 2017].
- COUTO-CORRÊA, F. *Gestión de datos científicos*. Barcelona: UOC, 2016.
- DALLMEIER-TIESSEN, S. et al. *Compilation of results on drivers and barriers and new opportunities* [en línea], 2012. Disponible en: <<http://epic.awi.de/31394>>. [Consulta: 13 de enero de 2017].
- DATA CITE. *DataCite metadata schema 3.1 XML schema* [web], 2014. Disponible en: <http://dx.doi.org/10.5438/0010>.
- DOORN, P. y DILLO, I. (2016). *FAIR data in trustworthy data repositories* [en línea]. Disponible en: <[http://www.slideshare.net/EUDAT/fair-data-in-trustworthy-data-repositories-webinar-1213-december-2016?qid=b369efe3-48d4-4442-a548-77d9be8ca9be&v=&b=&from\\_search=1](http://www.slideshare.net/EUDAT/fair-data-in-trustworthy-data-repositories-webinar-1213-december-2016?qid=b369efe3-48d4-4442-a548-77d9be8ca9be&v=&b=&from_search=1)>. [Consulta: 20 de enero de 2017].
- ELGER, K. et al. Open research data, data portals and data publication. An introduction to the data curation landscape [en línea]. *Polarforschung*, 2016, vol. 85, n° 2, p. 119-133. Disponible en: <http://dx.doi.org/10.2312/polfor.2016.009>.
- ERWAY, R. *Starting the conversation: University-wide research data management policy* [en línea]. Dublin, Ohio: OCLC Research, 2013. Disponible en: <<http://www.oclc.org/research/publications/library/2013/2013-08r.html>>. [Consulta: 13 de enero de 2017].
- ESPAÑA. Ley 14/2011, de 1 de junio, de la ciencia, la tecnología y la innovación [en línea]. BOE, n° 131, p. 54387-54455. CVE: BOE-A-2011-9617. Disponible en: <<http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf>>. [Consulta: 13 de enero de 2017].
- ESPAÑA. Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público [en línea]. BOE, n° 164, 10 de julio. CVE: BOE-A-2015-7731. Disponible en: <[https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2015-7731](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-7731)>. [Consulta: 13 de enero de 2017].
- FARQUHAR, A. y BRASE, J. Data identification and citation. The key to unlocking the promise of data sharing and reuse [en línea]. *D-LIB magazine*, 2014, vol. 20, n° 1/2. Disponible en: <http://dx.doi.org/10.1045/january2014-farquhar>.
- FEAR, K. The impact of data reuse: a pilot study of 5 measures [en línea]. En: *Research Data Access & Preservation Summit 2013*, Baltimore, MD, april 4, 2013. Disponible en: <[http://www.slideshare.net/asist\\_org/kfear-rdap](http://www.slideshare.net/asist_org/kfear-rdap)>. [Consulta: 13 de enero de 2017].
- FECHER, B. et al. A reputation economy: Results from an empirical survey on academic data sharing [en línea]. *DIW Discussion paper*, Berlin: DIW Berlin, German Institute for Economic Research, 2015. Disponible en: <<https://www.econstor.eu/handle/10419/107687>>. [Consulta: 13 de enero de 2017].
- FERRER-SAPENA, A. y ALEIXANDRE-BENAVENT, R. La opinión y los hábitos de los investigadores sobre los datos de investigación [en línea]. En: *Seminario Gestión de Datos de Investigación*. Universidad de Barcelona, 19 enero 2016. Disponible en: <[http://bd.ub.edu/grups/ccd/seminari\\_maredata](http://bd.ub.edu/grups/ccd/seminari_maredata)> y <<http://www.slideshare.net/maredata/presentacion-barcelona-3rafaelantoniv2-57839613>>. [Consulta: 13 de enero de 2017].

- FERRER-SAPENA, A.; PESET, F. y ALEIXANDRE-BENAVENT, R. Acceso a los datos públicos y su reutilización: open data y open government [en línea]. *El profesional de la información*, 2011, vol. 20, nº 3, p. 260-269. Disponible en: <http://dx.doi.org/10.3145/epi.2011.may.03>.
- FERRER-SAPENA, A. y SÁNCHEZ-PÉREZ, E.A. Open data, big data: ¿hacia dónde nos dirigimos? [en línea]. *Anuario ThinkEPI*, 2013, vol. 7, p. 150-156. Disponible en: <http://recyt.fecyt.es/index.php/ThinkEPI/article/view/30351>. [Consulta: 13 de enero de 2017].
- FERRER-SAPENA, A. et al. Cómo analizar el impacto de los datos de investigación con métricas: modelos y servicios [en línea]. *El profesional de la información*, 2016, vol. 25, nº 4. Disponible en: <http://dx.doi.org/10.3145/epi.2016.jul.13>.
- FILIPPOV, S. Mapping text and data mining in academic and research communities in Europe [en línea]. *Lisbon council. Think tank for the 21<sup>st</sup> century*, n. 16, 2014. Disponible en: <http://www.lisboncouncil.net/publication/publication/109-mapping-text-and-data-mining-in-academic-and-research-communities-in-europe.html>. [Consulta: 13 de enero de 2017].
- FORCE, M.M. y ROBINSON, N.J. Encouraging data citation and discovery with the Data Citation Index [en línea]. *Journal of computer-aided molecular design*, 2014, vol. 28, nº 10, p. 1043-1048. Disponible en: <http://dx.doi.org/10.1007/s10822-014-9768-5>.
- GARCÍA-GARCÍA, A.; LÓPEZ-BORRULL, A. y PESET, F. Análisis de *data journals*: la eclosión de nuevas revistas especializadas en datos [en línea]. *El profesional de la información*, 2015, vol. 24, nº 6, p. 845-854. Disponible en: <http://dx.doi.org/10.3145/epi.2015.nov.17>.
- GARCÍA-GARCÍA, A. y PESET, F. El camino incierto de la gestión de los datos [en línea]. *Blok de BiD*, 26 febrero 2014. Disponible en: <http://www.ub.edu/blokdebid/es/content/el-camino-incierto-de-la-gestion-de-los-datos>. [Consulta: 13 de enero de 2017].
- GARCÍA-GARCÍA, A. et al. *ODiSEA*: International registry on research data [en línea]. *BiD: textos universitaris de biblioteconomia i documentació*, 2012, nº 29. Disponible en: <http://www.ub.edu/bid/29/garcia2.htm>. [Consultado: 13 de enero de 2017].
- GONZÁLEZ, L.M. et al. Gestión de datos de investigación: infraestructuras para su difusión [en línea]. *El profesional de la información*, 2013, vol. 22, nº 5, p. 415-423. Disponible en: <http://dx.doi.org/10.3145/epi.2013.sep.06>.
- HALBERT, M. Prospects for research data management. En: Council on Library and Information Resources [en línea]. *Research data management: Principles, practices, and prospects*. Washington, D.C.: CLIR, 2013. Disponible en: <https://www.clir.org/pubs/reports/pub160/pub160.pdf>. [Consulta: 13 de enero de 2017].
- HE, L. y NAHAR, V. Reuse of scientific data in academic publications [en línea]. *Aslib journal of information management*, 2016, vol. 68, nº 4, p. 478-494. Disponible en: <http://dx.doi.org/10.1108/ajim-01-2016-0008>.
- JOHNSON, R.; PINFIELD, S. y FOSCI, M. Business process costs of implementing 'gold' and 'green' open access in institutional and national contexts [en línea]. *Journal of the Association for Information Science and Technology*, 2015, vol. 67, nº 9, p. 2283-2295. Disponible en: <http://dx.doi.org/10.1002/asi.23545>.
- KERALIS, S.D.C. et al. Research data management in policy and practice: The dataRes project. En: Council on Library and Information Resources [en línea]. *Research data management: Principles, practices, and prospects*. Washington, D.C.: CLIR, 2013. Disponible en: <https://www.clir.org/pubs/reports/pub160/pub160.pdf>. [Consulta: 13 de enero de 2017].
- KONKIEL, S. Tracking citations and altmetrics for research data: Challenges and opportunities [en línea]. *Bulletin of the American Society for Information Science and Technology*, 2013, vol. 39, nº 6, p. 27-32. Disponible en: [http://eprints.rclis.org/20028/1/AugSep13\\_Konkiel.pdf](http://eprints.rclis.org/20028/1/AugSep13_Konkiel.pdf). [Consulta: 13 de enero de 2017].
- LARIVIÈRE, V.; HAUSTEIN, S. y MONGEON, P. The oligopoly of academic publishers in the digital era [en línea]. *PLoS ONE*, 2015, vol. 10, nº 6. Disponible en: <http://dx.doi.org/10.1371/journal.pone.0127502>.
- LAVOIE, B.; CHILDRESS, E. et al. *The Evolving Scholarly Record* [en línea]. Ohio, OCLC, 2014. Disponible en: <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-evolving-scholarly-record-2014.pdf>. [Consulta: 20 de enero de 2017].
- MARTÍNEZ-URIBE, L. y MACDONALD, S. Un nuevo cometido para los bibliotecarios académicos: data curation [en línea]. *El profesional de la información*, 2008, vol. 17, nº 3, p. 273-280. Disponible en: <http://dx.doi.org/10.3145/epi.2008.may.03>.
- MARTINSEN, D.P. Supplemental journal article materials. En: Xiao, N.; McEwen, Leah-Rae (eds.). *Special issues in data management; ACS Symposium Series* [en línea]. Washington, DC: American Chemical Society, 2012. Disponible en: <http://dx.doi.org/10.1021/bk-2012-1110.ch003>.
- MENDEZ, E. "Cool" metadata for FAIR data [en línea]. Disponible en: <http://www.slideshare.net/ResearchDataAlliance/cool-metadata-for-fair-data>. [Consulta: 30 de enero de 2017].
- MOEDAS, C. Open innovation, open science, open to the world [en línea]. En: Comisión europea. *A new start for Europe: Opening up to an ERA of Innovation Conference*, SPEECH/15/5243, 2015. Disponible en: [http://europa.eu/rapid/press-release\\_SPEECH-15-5243\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-15-5243_en.htm) y <http://ec.europa.eu/research/conferences/2015/era-of-innovation/index.cfm>. [Consulta: 13 de enero de 2017].

- MOSS, E. Viable data citation: Expanding the impact of social science research [en línea]. En: *RDAP13. Panel on Data Citation and Altimetrics*, Abril 5, 2013. Disponible en: <[http://es.slideshare.net/asist\\_org/rdap13-moss](http://es.slideshare.net/asist_org/rdap13-moss)>. [Consulta: 13 de enero de 2017].
- MURRAY-RUST, P.; MOLLOY, J.C. y CABELL, D. Open content mining. En: Moore, Samuel (ed.). *Issues in open research data* [en línea]. Mountain View, CA: Ubiquity Press, 2014, p. 11-31. Disponible en: <<http://www.ubiquitypress.com/site/books/detail/12/issues-in-open-research-data>>. [Consulta: 13 de enero de 2017].
- NATIONAL RESEARCH COUNCIL. *Bits of power: Issues in global access to scientific data* [en línea]. Washington, D.C.: NATIONAL ACADEMY PRESS, 1997. Disponible en: <<https://www.nap.edu/read/5504/chapter/1>>. [Consulta: 13 de enero de 2017].
- NATIONAL SCIENCE FOUNDATION. *Dissemination and sharing of research results*. National Science Foundation Data Sharing Policy [web], 2010. Disponible en: <<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>>. [Consulta: 13 de enero de 2017].
- NIH. *Final NIH Statement on sharing research data*. NOT-OD-03-032 [en línea], February 26. National Institutes of Health, 2003. Disponible en: <<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>>. [Consulta: 13 de enero de 2017].
- NINA-ALCOCER, V.; BLASCO-GIL, Y. y PESET, F. *Datasharing: guía práctica para compartir datos de investigación* [en línea]. *El profesional de la información*, 2013, vol. 22, nº 6, p. 562-568. Disponible en: <http://dx.doi.org/10.3145/epi.2013.nov.09>.
- NISO. *Recommended practices for online supplemental journal article materials*. NISO RP-15-2013 [en línea]. Baltimore, MD: National Information Standards Organization, 2013. Disponible en: <<http://www.niso.org/publications/rp/rp-15-2013>>. [Consulta: 13 de enero de 2017].
- NOSEK, B. *et al.* Promoting an open research culture. *Science* [en línea], 2015, vol. 348, nº 6242, p. 1422-1425. Disponible en: <http://dx.doi.org/10.1126/science.aab2374>.
- OECD. *Principles and guidelines for access to research data from public funding* [en línea]. París: OECD Publishing, 2007. Disponible en: <<https://www.oecd.org/sti/sci-tech/38500813.pdf>>. [Consulta: 13 de enero de 2017].
- PAMPEL, H. *et al.* Making research data repositories visible: The re3data.org registry [en línea]. *PLoS ONE*, 2013, vol. 8, nº 11. Disponible en: <http://dx.doi.org/10.1371/journal.pone.0078080>.
- PESET, F. y FERNÁNDEZ-LÓPEZ, A.L. Carencias informativas de los datos abiertos en España [en línea]. *Anuario ThinkEPI*, 2014, vol. 8, p. 318-321. Disponible en: <<http://recyt.fecyt.es/index.php/ThinkEPI/article/view/29599>>. [Consulta: 13 de enero de 2017].
- PESET, F. *et al.* Research data management in Spain: results from a *Datasea* project survey [en línea]. En: *10th International Digital Curation Conference*. Londres, 9-12 febrero, 2015. Disponible en: <[http://www.dcc.ac.uk/sites/default/files/documents/IDCC15/205\\_RDMSpain-DATASEA.pdf](http://www.dcc.ac.uk/sites/default/files/documents/IDCC15/205_RDMSpain-DATASEA.pdf)>. [Consulta: 13 de enero de 2017].
- PESET, F. y GONZÁLEZ, L.M. *Ciencia abierta y gestión de datos de investigación*. Gijón: TREA, 2016.
- PETERS, I. *et al.* Research data explored: an extended analysis of citations and altmetrics [en línea]. *Scientometrics*, 2016, vol. 107, nº 2, p. 723-744. Disponible en <http://dx.doi.org/10.1007/s11192-016-1887-4>.
- PIWOWAR, H.A. *et al.* Towards a data sharing culture: recommendations for leadership from academic health centers [en línea]. *PLoS Med*, 2008, vol. 5, nº 9. Disponible en: <http://doi.org/10.1371/journal.pmed.0050183>.
- PIWOWAR, H.A. y CHAPMAN, W. A review of journal policies for sharing research data [en línea]. En: *Proceedings ELPUB 2008. Conference on Electronic Publishing. Toronto, June*. Disponible en: <[http://elpub.scix.net/data/works/att/001\\_elpub2008.content.pdf](http://elpub.scix.net/data/works/att/001_elpub2008.content.pdf)>. [Consulta: 13 de enero de 2017].
- PIWOWAR, H.A. y CHAPMAN, W. Public sharing of research datasets: A pilot study of associations [en línea]. *Journal of informetrics*, 2010, vol. 4, nº 2, p. 148-156. Disponible en: <http://doi.org/10.1016/j.joi.2009.11.010>.
- PIWOWAR, H.A.; DAY, R.S. y FRIDSMA, D.B. Sharing detailed research data is associated with increased citation rate [en línea]. *PLoS ONE*, 21 marzo 2007. Disponible en: <http://dx.doi.org/10.1371/journal.pone.0000308>.
- PIWOWAR, H.A. y VISION, T.J. Data reuse and the open data citation advantage [en línea]. *PeerJ*, 2013, 1:e175. Disponible en: <http://doi.org/10.7717/peerj.175>.
- PORTER, J.H. A brief history of data sharing in the US long term ecological research network [en línea]. *The bulletin of the Ecological Society of America*, 2010, vol. 91, nº 1, p. 14-20. Disponible en: <<http://onlinelibrary.wiley.com/doi/10.1890/0012-9623-91.1.14/full>>. [Consulta: 13 de enero de 2017].
- RANI, M. y BUCKLEY, B.S. Systematic archiving and access to health research data: rationale, current status and way forward [en línea]. *Bulletin of the world health organization*, 2012, vol. 90, nº 12, p. 932-939. Disponible en: <<http://www.who.int/bulletin/volumes/90/12/12-105908/en/>>. [Consulta: 13 de enero de 2017].
- RE3DATA. Re3data.org reaches a milestone & begins offering badges. *Re3data.org* [web], 2016. Disponible en: <<http://www.re3data.org/2016/04/re3data-org-reaches-a-milestone-begins-offering-badges>>. [Consulta: 13 de enero de 2017].

- ROBINSON-GARCÍA, N.; JIMÉNEZ-CONTRERAS, E. y TORRES-SALINAS, D. Analyzing data citation practices using the data citation index [en línea]. *Journal of the Association for Information Science and Technology*, 2015. Disponible en: <http://dx.doi.org/10.1002/asi.23529>.
- ROYAL SOCIETY. *Science as an open enterprise* [en línea]. Londres: The Royal Society; Science Policy Centre, 2012. Disponible en: [https://royalsociety.org/~media/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf). [Consulta: 13 de enero de 2017].
- Thomson Reuters. *Unlocking the value of research data* [en línea], 2013. Disponible en: <http://researchanalytics.thomsonreuters.com/m/pdfs/1003903-1.pdf>. [Consulta: 13 de enero de 2017].
- TORRES-SALINAS, D. Compartir datos (*data sharing*) en ciencia: contexto de una oportunidad [en línea]. *Anuario ThinkEPI*, 2010, vol. 4, p. 258-261. Disponible en: <http://recyt.fecyt.es/index.php/ThinkEPI/article/view/31270>. [Consulta: 13 de enero de 2017].
- TORRES-SALINAS, D.; MARTIN-MARTIN, A. y FUENTE-GUTIERREZ, E. Analysis of the coverage of the data citation index. Thomson Reuters: disciplines, document types and repositories [en línea]. *Revista española de Documentación Científica*, 2014, vol. 37, nº 1. Disponible en: <http://dx.doi.org/10.3989/redc.2014.1.1114>.
- TORRES-SALINAS, D.; ROBINSON-GARCÍA, N. y CABEZAS-CLAVIJO, A. Compartir los datos de investigación: introducción al *data sharing* [en línea]. *El profesional de la información*, 2012, vol. 21, nº 2, p. 173-184. Disponible en: <http://dx.doi.org/10.3145/epi.2012.mar.08>.
- UNIÓN EUROPEA. *Comunicación de la comisión al parlamento europeo, al consejo, al comité económico y social europeo y al comité de las regiones. Hacia una economía de los datos próspera (COM(2014) 442 final)*. 2014 [en línea]. Disponible en: <http://ec.europa.eu/transparency/regdoc/rep/1/2014/ES/1-2014-442-ES-F1-1.Pdf>. [Consulta: 13 de enero de 2017].
- UNIÓN EUROPEA. *H2020 Programme. Guidelines on data management in horizon 2020* [en línea]. Versión 3.0. 26 July 2016. 2016a. Disponible en: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). [Consulta: 13 de enero de 2017].
- UNIÓN EUROPEA. *European cloud initiative to give Europe a global lead in the data-driven economy* [en línea], 2016b. Disponible en: [http://europa.eu/rapid/press-release\\_IP-16-1408\\_en.htm](http://europa.eu/rapid/press-release_IP-16-1408_en.htm). [Consulta: 13 de enero de 2017].
- VLAEMINCK, S. y WAGNER, G.G. On the role of research data centres in the management of publication-related research data [en línea]. *Liber quarterly*, 2014, vol. 23, nº 4, p. 336-357. Disponible en: <http://doi.org/10.18352/lq.9356>.
- WALLIS, J.C.; ROLANDO, E. y BORGMAN, C.L. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology [en línea]. *PLoS ONE*, 2013, vol. 8, nº 7, p. e67332. Disponible en: <http://doi.org/10.1371/journal.pone.0067332>.
- WARE, M. y MABE, M. The STM report: An overview of scientific and scholarly journal publishing. Celebrating the 350th anniversary of journal publishing [en línea]. 4th ed. The Hague: International Association of Scientific, Technical and Medical Publishers, 2015. Disponible en: [http://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf). [Consulta: 13 de enero de 2017].